Flexible Spatial Dependence Structures for Unordered Multinomial Choice Models: Formulation and Application to Teenagers' Activity Participation

Ipek N. Sener

Texas Transportation Institute
Texas A&M University System
1106 Clayton Lane, Suite 300E, Austin, TX, 78723
Phone: (512) 467-0952, Fax: (512) 467-8971
Email: i-sener@ttimail.tamu.edu

and

Chandra R. Bhat*

The University of Texas at Austin
Department of Civil, Architectural & Environmental Engineering
1 University Station, C1761, Austin, TX 78712-0278
Phone: (512) 471-4535, Fax: (512) 475-8744
Email: bhat@mail.utexas.edu

*corresponding author

Original: July 29, 2010
Revised: April 27, 2011
Second Revision: August 5, 2011

ABSTRACT

The current paper proposes an approach to accommodate flexible spatial dependency structures in discrete choice models in general, and in unordered multinomial choice models in particular. The approach is applied to examine teenagers' participation in social and recreational activity episodes, a subject of considerable interest in the transportation, sociology, psychology, and adolescence development fields. The sample for the analysis is drawn from the 2000 San Francisco Bay Area Travel Survey (BATS) as well as other supplementary data sources. The analysis considers the effects of a variety of built environment and demographic variables on teenagers' activity behavior. In addition, spatial dependence effects (due to common unobserved residential neighborhood characteristics as well as diffusion/interaction effects) are accommodated. The variable effects indicate that parents' physical activity participation constitutes the most important factor influencing teenagers' physical activity participation levels, In addition, part-time student status, gender, and seasonal effects are also important determinants of teenagers' social-recreational activity participation. The analysis also finds strong spatial correlation effects in teenagers' activity participation behaviors.

1. INTRODUCTION

Spatial dependence is inherent in many aspects of human decision-making, with the choice decisions of one individual being affected by those of other individuals who are proximal in space. This inter-relationship in decision-making may be a consequence of several reasons, including diffusion effects, social interaction effects, or unobserved location-related influences (see Jones and Bullen, 1994, and Miller, 1999). The importance of such spatial dependence effects has been recognized for several decades now in a variety of disciplines, including geography, urban planning, economics, political science, and transportation to name just a few (see Páez, 2007 and Franzese and Hays, 2008 for recent reviews). However, much of the work explicitly recognizing such dependence in modeling human decision-making directly, or as an aggregation of decisions across several individuals residing in a "neighborhood", has been confined to situations where the variable of interest is continuous (see, for instance, Cho and Rudolph, 2007, Boarnet *et al.*, 2005, Messner and Anselin, 2004). On the other hand, many choice decisions in the context of activity-travel analysis and several other fields are inherently discrete, and can be strongly influenced by spatial considerations. In this regard, the current study contributes to the area of spatial analysis in discrete choice modeling by developing a flexible econometric modeling approach that accounts for spatial dependence in an unordered multinomial choice model setting.

The application context of the study is teenagers' participation in social and recreational episodes. Specifically, a choice model is used to model teenagers' participation in social, physically inactive recreation, and physically active recreation episodes (the precise definitions of these activity purposes are provided later). A flexible spatial error dependence in participation propensities in these activity purposes is generated across teenagers based on the proximity of their residences. Such dependencies may be the result of unobserved residential urban form factors (such as good bicycle and walk path continuity) that may increase participation tendencies in specific activities, or diffusion and social interaction effects between proximally located teenagers so that unobserved lifestyle perspectives (such as physically active lifestyle attitudes) that affect activity participation decisions become correlated. We accommodate spatial error correlation through a copula structure that does not pre-impose any dependence structure. For instance, for a given (say positive) spatial correlation, the traditional multivariate normal dependence structure imposes the assumption that proximally located teenagers may have a simultaneously low propensity for physically active recreational participation or a simultaneously high propensity for physically active recreational

participation. However, the multivariate normal dependence structure does not allow asymmetric dependence structures, such as would be the case if proximally located teenagers have a simultaneously high propensity for physically active recreational participation but not necessarily a simultaneously low propensity for physical activity participation. That is, unobserved factors that increase physical activity propensity may diffuse more among teenagers than unobserved factors that decrease physical activity propensity. Such a spatial correlation pattern can only be reflected through the use of a copula dependence structure that has strong right tail dependence (strong correlation at high values) but weak left tail dependence (weak correlation at low values). Our approach allows the comparison of such an asymmetric dependency structure with the symmetric multivariate normal (or Gaussian) dependency structure.

Teenagers' participation in social and recreational activity episodes, the application focus of this paper, is an important area of study in several fields, including child development, public health, and transportation. In the child development field, many studies have established the positive role that out-of-home social-recreational activity participation plays in children's self-development in the context of social skills, self-esteem, identity exploration, sense of responsibility, and understanding of fairness concepts (see, for instance, Darling, 2005, and Campbell, 2007). This is particularly so during adolescence due to the rapid emotional and physical personality developments at this life stage (Fredricks and Eccles, 2008). In fact, as indicated by Sanchez-Samper and Knight (2009), "adolescence is a time of physical, emotional, and psychological maturation as well as a period of searching for independence and experimentation". However, along with the potentially substantial mental/physical growth and independence that adolescents experience, this is also a period when individuals are prone to gravitate toward health-risky behaviors such as drug use, tobacco use, and unprotected sex (see Tiggemann, 2001, and Lerner and Steinberg, 2004). Such behaviors can be controlled and reduced by motivating adolescents to participate in social-recreational activities that provide a vehicle to develop healthy and communicative relationships with peers and adults (see Eccles and Gootman, 2002). Focusing on the factors that influence participation in social-recreational activities as a way to reduce health-risky behaviors among adolescents is also consistent with a "positive youth development" (PYD) paradigm approach to address challenges during the adolescence period (as opposed to much child development research that focuses almost exclusively on intervention programs to restrain risky behaviors; see Larson, 2000, who initiated research on the PYD paradigm).

Teenagers' participation in social-recreational activities has also been an important area of research in the public health field. In addition to the mental health issues that overlap with the child development literature, the participation of teenagers in physically active recreational pursuits has interested public health researchers for some time now. The current paper contributes to this research area, particularly because we differentiate between physically active and physically inactive recreation activities within the category of recreational activities. As is now well established in the public health literature, sedentary (or physically inactive) life styles are associated with obesity, heart disease, diabetes, high blood pressure, and several forms of cancer and mental health diseases (see, for instance, Nelson and Gordon-Larsen, 2006, Centers for Disease Control and Prevention (CDC), 2006, and Ornelas *et al.*, 2007). On the other hand, physical activity increases cardiovascular fitness, enhances agility and strength, reduces the need for medical attention, and contributes to improved mental health, and decreases depression and anxiety.[1] But despite the negative physical health consequences of sedentary lifestyles and the positive benefits of an active lifestyle, about a third of teenagers do not engage in adequate physical activity for health, and this low-level of physical activity participation is particularly acute among older teenagers and teenage girls (CDC, 2010).

The study of teenagers' out-of-home social-recreational activity participation is not just relevant to the child development and public health fields. Analyzing and modeling activity-travel patterns of children, and teenagers in particular, has started to attract increasing attention in the activity-based travel demand modeling field since children's/teenagers' activities inherently influence, and are influenced by, adults' activity-travel patterns (see, for instance, McDonald, 2005, Sener *et al.*, 2008, Stefan and Hunt, 2006). Adults (especially parents) spend a considerable amount of time escorting children and teenagers to out-of-home activities, and participating with children in joint social-recreational activities (Reisner, 2003, McGuckin and Nakamoto, 2004, and Sener and Bhat, 2007). The weekday focus of the current study is particularly important because of the increased amount of adults' activity episodes and trips attributable to children's/teenagers' after-school social-recreational activity participation (Reisner, 2003). Indeed, studies in the literature have pointed out that children as young as 6-8 years start developing their own identities and

---

[1] The statements here are not intended to imply that all sedentary activities are unhealthy activities. As discussed earlier, participation in social-recreational activities (regardless of physical activity levels) can be helpful in a youth's overall personal development.

individualities, and social needs (see Stefan and Hunt, 2006, CDC, 2005, Eccles, 1999). They then interact with their parents and other adults to facilitate these activity-travel needs. Also, the consideration of children's activity-travel patterns is important in its own right because these patterns contribute directly to travel demand. For instance, using data from the 2002 Child Development Supplement to the Panel Study of Income Dynamics, Paleti *et al.* (2011) found that a significant percentage of teenagers (about 35% in the US) do not return home immediately after school, and the majority of activities pursued by these teenagers at the out-of-home location are social-recreational in nature.

The rest of this paper is structured as follows. The next section presents the structure of the copula-based spatial multinomial unordered response model and discusses the (composite marginal likelihood) estimation approach employed in the current paper. Section 3 presents description of the data source and sample formation procedures used in the empirical context of our study. Section 4 presents the empirical analysis results. The final section summarizes the important findings and concludes the paper.

2. MODEL FORMULATION

2.1. Copula-based Spatial Unordered Response Model Structure

Let $\widetilde{u}_{qi}^{*}$ be the indirect (latent) utility of the qth observational unit for the ith alternative ($q = 1, 2, \ldots, Q$; $i = 1, 2, \ldots, I$).[2] Let $\widetilde{u}_{qi}^{*}$ be written in the usual way as a linear combination of a deterministic component V_{qi} and a stochastic component ε_{qi}. The deterministic component is assumed to be linear-in-parameters; $V_{qi} = \beta' x_{qi}$, where x_{qi} is a vector of exogenous variables and β is a corresponding coefficient vector. The error terms ε_{qi} are assumed to be type I extreme value (Gumbel) distributed with a scale parameter of σ_q (this allows for heteroscedasticity across observation units).

Let D_{qi} be a dummy variable indicator that takes the value of 1 if q selects i and 0 otherwise. Since the alternatives $i = 1, 2, \ldots, I$ are collectively exhaustive, the values of D_{q1}, D_{q2}, $\ldots, D_{qI-1}$

---

[2] We will use general notation in the presentation of the model formulation to emphasize the generality of the formulation for multinomial discrete choice analysis. In the specific empirical context of the current paper, q is the index for

suffice to characterize the chosen alternative for q (that is, if $D_{qi} = 0$ for $i = 1, 2, \ldots, I-1$, it automatically implies that individual q chooses alternative I). Then, in the usual tradition of random utility maximization (RUM), we can write:

$$D_{qi} = 1 \text{ for } i = 1, 2, \ldots, I-1 \quad \text{if and only if } \tilde{u}_{qi}^* > \max_{\substack{k=1,2,\ldots,I \\ k \neq i}} \tilde{u}_{qk}^* \tag{1}$$

$$D_{qI} = 1 \text{ for the last alternative } I \text{ if and only if } \tilde{u}_{qi}^* < \max_{\substack{k=1,2,\ldots,I \\ k \neq i}} \tilde{u}_{qk}^* \text{ for all } i = 1, 2, \ldots, I-1$$

Next, define a random variable as follows:

$$h_{qi} = \left\{ \max_{\substack{k=1,2,\ldots,I \\ k \neq i}} \tilde{u}_{qk}^* \right\} - \varepsilon_{qi} \, (i = 1, 2, \ldots, I-1)$$

Then, the equation system in Equation (1) can be written as:

$$D_{qi} = 1 \text{ for } i = 1, 2, \ldots, I-1 \text{ if and only if } h_{qi} < V_{qi} \tag{2}$$

$$D_{qI} = 1 \text{ for alternative } I \quad \text{if and only if } h_{qi} > V_{qi} \text{ for all } i \neq I$$

Let $H_{\varepsilon_{q1}, \varepsilon_{q2}, \ldots, \varepsilon_{qI}}(z_{q1}, z_{q2}, \ldots, z_{qI})$ be the multivariate cumulative distribution of the alternative error terms ε_{q1}, ε_{q2}, …, ε_{qI}. In the case when the error terms are independent and identically distributed (IID) across alternatives with each error term being Gumbel distributed with scale σ_q, this multivariate distribution is:

$$H_{\varepsilon_{q1}, \varepsilon_{q2}, \ldots, \varepsilon_{qI}}(z_{q1}, z_{q2}, \ldots, z_{qI}) = \exp\left\{ -\sum_{i=1}^{I} \exp\left[-\frac{z_{qi}}{\sigma_q} \right] \right\}. \tag{3}$$

With the IID error distribution across alternatives for the error terms, the implied marginal distribution of h_{qi} ($i = 1, 2, \ldots, I-1$) is:

---

teenagers and i is the index for the type of social-recreational activity chosen for participation at each episode choice instance.

[3] On the other hand, if the error terms are Generalized Extreme Value (GEV) distributed across alternatives with identical scale parameters, this equation takes the familiar GEV form. In the rest of this section, we will consider the error terms to be IID across alternatives for ease in presentation, though extension to the GEV structure is straightforward. In fact, in the empirical analysis, we explore nested logit models (a form of the GEV structure).

$$P(h_{qi} < s) = G_{qi}(s) = \frac{e^{s|\sigma_q}}{e^{s|\sigma_q} + \sum_{k \neq i} e^{(V_{qk}|\sigma_q)}} = \frac{1}{1 + e^{-\left[\frac{s - \sigma_q \ln \sum_{k \neq 1} e^{(V_{qk}|\sigma_q)}}{\sigma_q}\right]}}, \quad (4)$$

which is logistic distributed. Further, if there is no spatial correlation in the error terms ε_{qi} across observation units q, then the probability above $P(h_{qi} < s)$ is independent of the probability $P(h_{q'i} < s)$ for all $q' \neq q$ (because of the construction of the h_{qi} variable). Thus, from Equation (2), we can write the probability of choice of observation unit q for alternative i ($i = 1, 2, \ldots, I - 1$) as:

$$P(D_{qi} = 1) = P(h_{qi} < V_{qi}) = G_{qi}(V_{qi}) = \frac{e^{V_{qi}|\sigma_q}}{e^{V_{qi}|\sigma_q} + \sum_{k \neq i} e^{(V_{qk}|\sigma_q)}} = \frac{e^{V_{qi}|\sigma_q}}{\sum_k e^{(V_{qk}|\sigma_q)}}.$$

This, of course, corresponds to the case of the simple heteroscedastic multinomial logit (IHMNL) model for each observation unit. Further, if the scale σ_q is identical across all observation units, the result is the independent multinomial logit (IMNL) model.

The situation is more difficult when the error terms ε_{qi} are dependent across observation units q for each i (that is, $\text{cov}(\varepsilon_{qi}, \varepsilon_{q'i}) \neq 0$, though we will maintain that $\text{cov}(\varepsilon_{qi}, \varepsilon_{q'i'}) = 0$). The dependence in errors across observation units for each alternative can arise through spatial dependency effects, as we postulated earlier in this paper. In this case, the random terms h_{qi} ($q = 1$, 2, …, Q, $i = 1, 2, \ldots, I - 1$) with the pre-specified parametric marginal distributions $G_{qi}(.)$ are no more independent across observations q for each i. In the current paper, we tie the h_{qi} random terms together across observations q for each i ($i = 1, 2, \ldots, I - 1$) using a copula, which is a device or function that generates a stochastic dependence relationship (*i.e.*, a multivariate distribution) among random variables with pre-specified marginals (see Bhat and Eluru, 2009 for a detailed discussion of copulas). The power of the copula approach is that it disassociates the marginal distributions of random variables from the dependence structure. Let C_θ be the *Q*-dimensional copula considered for each alternative *i*, with θ being a parametric vector of the copula referred to as the dependence

parameter vector.[4] Then by Sklar's (1973) theorem, a joint Q-dimensional distribution function for the random variables h_{qi} $(q = 1, 2, \ldots, Q)$ with the continuous marginal distribution functions $G_{qi}(s)$ can be generated as follows:

$$F_i(s_1, s_2, \ldots, s_Q) = \Pr(h_{1i} < s_1, h_{2i} < s_2, \ldots, h_{Qi} < s_Q = C_\theta[G_{1i}(s_1), G_{2i}(s_2), \ldots, G_{Qi}(s_Q)].[5] \tag{5}$$

Several different multivariate copulas exist in the literature, though there are only a limited number of these that can allow for differential dependence intensities among pairs of variables. In the context of spatial dependence, one expects such differential dependence intensities between observational units q based on spatial proximity. Three types of flexible multivariate copulas that are well suited for spatial dependence are the Gaussian, Farlie-Gumbel-Morgenstein (FGM), and the Generalized Gumbel Copula recently proposed by Bhat (2009). Of these, the first two copulas are radically symmetric and assume the property of asymptotic independence. The third copula, on the other hand, allows asymmetric and extreme tail dependence (*i.e.*, the dependence is higher in the right tail than in the left tail). It also allows only for positive dependence. To write these copula forms, consider the Q-dimensional copula C_θ of uniformly distributed random variables $U_1, U_2, U_3,$ $\ldots, U_K$ with support contained in $[0,1]^Q$. Then, the three copula structures are as follows.

Multivariate Gaussian Copula

$$\begin{aligned} C_\theta(u_{1i}, u_{2i}, \ldots, u_{Qi}) &= \Pr(U_{1i} < u_{1i}, U_{2i} < u_{2i}, \cdots U_{Qi} < u_{qi}) \\ &= \Phi_Q(\Phi^{-1}(u_{1i}), \Phi^{-1}(u_{2i}), \ldots, \Phi^{-1}(u_{Qi}), \theta), \end{aligned} \tag{6}$$

where Φ_Q is the Q-dimensional standard normal cumulative distribution function (CDF) with zero mean and a correlation matrix whose off-diagonal elements are captured in the vector θ, and $\Phi^{-1}(.)$ is the inverse of the univariate standard normal CDF. In the context of Equation (5), $u_{qi} = G_{qi}(s_i)$ for all $q = 1, 2, \ldots, Q$. The dependence structure in the Gaussian copula is radially symmetric about the center point. That is, for a given correlation, the level of dependence is equal in

---

[4] In fact, one can use different copulas to tie the h_{qi} terms across q for different alternatives i ($i = 1, 2, \ldots, I - 1$). In addition, the dependence parameter vector θ can vary across alternatives i. However, such flexibility also creates exchangeability problems, since the copulas (and the dependence vectors) estimated for each alternative i will not be independent of the decision of which alternative is considered as the last alternative I. Hence, we prefer the specification that restricts the copula (and the dependence vector) to be the same across alternatives i ($i = 1, 2, \ldots, I - 1$).

[5] If the random terms h_{qi} ($q = 1, 2, \ldots, Q$) are independent, then this equation collapses to:

the upper and lower tails. When all elements of θ are zero in the Gaussian copula, this implies independence among the uniform variates $U_1, U_2, U_3, \ldots, U_K$: $C(u_{1i}, u_{2i}, \ldots, u_{qi}, \ldots, u_{Qi}) = u_{1i} u_{2i} \ldots u_{qi}, \ldots u_{Qi}$.

Multivariate FGM Copula

$$C_\theta(u_{1i}, u_{2i}, \ldots, u_{Qi}) = \prod_{q=1}^{Q} u_{qi} \left[1 + \sum_{q=1}^{Q-1} \sum_{k=q+1}^{Q} \theta_{qk}(1 - u_{qi})(1 - u_{ki}) \right], \tag{7}$$

where θ_{qk} is the dependence parameter between U_q and U_k $(-1 \le \theta_{qk} \le 1)$, $\theta_{qk} = \theta_{kq}$ for all q and k. The FGM copula has a simple analytic form and allows for either negative or positive dependence. Like the Gaussian copula, it also imposes the assumptions of asymptotic independence and radial symmetry in dependence structure (see Bhat and Sener, 2009). When $\theta_{qk} = 0 \ \forall \ (q,k)$ pairs, we obtain the independence case.

Multivariate Generalized Gumbel Copula

$$C_\theta(u_{1i}, u_{2i}, \ldots, u_{Qi}) = \exp\left\{ -\sum_{q=1}^{Q-1} \sum_{k=q+1}^{Q} \left[(-\alpha_{q,qk} \ln u_{qi})^{1/\rho} + (-\alpha_{k,qk} \ln u_{ki})^{1/\rho} \right]^\rho \right\}, \tag{8}$$

$0 \le \alpha_{q,qk} \le 1$ for all q and k, and $\sum_j \alpha_{q,qk} = 1$ for all q, and $0 < \rho \le 1$.

The dependence parameter vector θ in this multivariate Generalized Gumbel (GG) copula includes the $\alpha_{q,qk}$ terms as well as the dissimilarity parameter ρ. This generalized version of the Gumbel copula, which is based on a cumulative multivariate extreme-value distribution, allows different dependence parameters between each variable pair U_q and U_k (see Bhat, 2009). Independence is achieved in the GG copula when $\rho = 1$ in Equation (8).

A couple of parameterizations are in order before proceeding to estimation. First, we parameterize σ_q in Equation (4) as $\sigma_q = g(\lambda' w_q) = \exp(\lambda' w_q)$ where w_q includes variables specific to pre-defined "neighborhoods" or other groupings of observational units and individual related factors, and λ is a corresponding coefficient vector to be estimated. If all elements of the vector λ

$F_i(s_1, s_2, \ldots, s_Q) = \Pr(h_{1i} < s_1) \times \Pr(h_{2i} < s_2) \times \ldots \Pr(h_{Qi} < s_{Qi}) = G_{1i}(s_1) \times G_{2i}(s_2) \times \ldots G_{Qi}(s_Q)$.

are zero, this implies no heteroscedasticity across observation units. Second, it is not possible to estimate a separate dependence term θ_{qk} for each pair of observational units q and k for the Gaussian and FGM copulas. So, we assume that the spatial process is isotropic (that is, $\theta_{qk} = \theta_{kq}$) and parameterize θ_{qk} for the Gaussian and FGM copulas as:

$$\theta_{qk} = \pm \frac{(e^\varsigma)'\tilde{s}_{qk}}{1+(e^\varsigma)'\tilde{s}_{qk}}, \tag{9}$$

where $\tilde{s}_{qk}$ is a vector of variables (taking on non-negative values) that influences the level of spatial dependence between observational units q and k choosing the same alternative, and ς is a corresponding set of parameters to be estimated. The functional form above ensures that $-1 \le \theta_{qk} \le 1$, as required in the FGM and Gaussian copulas (see Equations (6) and (7)). Further, in a spatial context, we expect observational units in close proximity to have similar preferences, because of which we impose the '+' sign in front of the expression in Equation (9). Note that the functional form of Equation (9) can accommodate various (and multiple) forms of spatial dependence through the appropriate consideration of variables in the vector $\tilde{s}_{qk}$ (see Bhat *et al.*, 2010 for a more detailed discussion of the reasons that motivate the functional form in Equation (9)). In particular, the dependence form nests the typical spatial dependence patterns used in the extant literature as special cases, including dependence based on (1) whether observational units are in the same "neighborhood" or in contiguous "neighborhoods" (obtained by including a single variable in the $\tilde{s}_{qk}$ vector that takes a value of 1 if q and k are in the same predefined "neighborhood" or in contiguous "neighborhoods", and 0 otherwise), (2) shared border length of the "neighborhood" of two observational units (obtained by having the border length or some functional transformation of the border length as the sole variable in the $\tilde{s}_{qk}$ vector, and (3) distance between observational units (obtained by including an appropriate representation of distance as the sole variable in the $\tilde{s}_{qk}$ vector).[6]

---

[6] Several functional forms of distance may be used, such as inverse of distance, square of inverse of distance, and distance "cliff" measures (the latter form essentially allows the spatial correlation between two teenagers to go to zero beyond a certain distance threshold). Also, the representation of distance may be in the form of time to travel or spatial distance, and may be measured as network distances or Euclidean distances ("crowfly" distances) or other measures of spatial separation.

For the generalized Gumbel copula of Equation (8), the dependence vector θ includes the $\alpha_{q,qk}$ terms as well as the ρ term. Since we cannot estimate a separate $\alpha_{q,qk}$ term for each pairing of observational units q and k choosing the same alternative, and also because we require that $0 \leq \alpha_{q,qk} \leq 1$ for all q and k choosing the same alternative and $\sum_k \alpha_{q,qk} = 1$ for all q, we use the following parameterization:

$$\alpha_{q,qk} = \frac{\exp(\delta \tilde{s}_{qk})}{\sum_l \exp(\delta \tilde{s}_{ql})} \tag{10}$$

The above form can include general forms of dependence based on the specification of $\tilde{s}_{qk}$. The dependence is strictly positive in the Generalized Gumbel Copula.

2.2. Estimation Approach

Without loss of generality, assume that the first Q_1 of the Q observational units in the data select alternative 1, the next $Q_1 + 1$ to Q_2 observational units select alternative 2, and so on. The likelihood function for the spatially correlated unordered MNL model may then be written using Equations (2) and (5) as:

$$L(\beta, \lambda, \varsigma \text{ or } \delta) = \Pr \left[\left\{ \begin{array}{l} \begin{cases} h_{11} < V_{11}, h_{21} < V_{21}, ..., h_{Q_1,1} < V_{Q_1,1}; \\ h_{Q_1+1,2} < V_{Q_1+1,2}, h_{Q_1+2,2} < V_{Q_1+2,2}, ..., h_{Q_2,2} < V_{Q_2,2}; \\ . \\ . \\ . \\ h_{Q_{(I-2)}+1,I-1} < V_{Q_{(I-2)}+1,I-1}, h_{Q_{(I-2)}+2,I-1} < V_{Q_{(I-2)}+2,I-1}, ..., h_{Q_{(I-1)},I-1} < V_{Q_{(I-1)},I-1} \end{cases} ; \\ \begin{cases} h_{Q_{(I-1)}+1,1} > V_{Q_{(I-1)}+1,1}, h_{Q_{(I-1)}+1,2} > V_{Q_{(I-1)}+1,2}, ..., h_{Q_{(I-1)}+1,I-1} > V_{Q_{(I-1)}+1,I-1}; \\ h_{Q_{(I-1)}+2,1} > V_{Q_{(I-1)}+2,1}, h_{Q_{(I-1)}+2,2} > V_{Q_{(I-1)}+2,2}, ..., h_{Q_{(I-1)}+2,I-1} > V_{Q_{(I-1)}+2,I-1}; \\ . \\ . \\ . \\ h_{Q,1} > V_{Q,1}, h_{Q,2} > V_{Q,2}, ..., h_{Q,I} > V_{Q,I} \end{cases} \end{array} \right\} \right] \tag{11}$$

Note that, as explained earlier, the alternatives $i = 1, 2, …, I$ are collectively exhaustive, and so the choice of alternative I for the last $Q_{(I-1)} + 1$ to Q observational units is equivalent to the non-choice

of the first $(I-1)$ alternatives. This is reflected in the multiple inequality conditions for each of these last observations within the second { } parenthesis in Equation (11). In the case of the extensively studied spatially correlated binary model, $I = 2$ and the likelihood collapses to the simpler form:

$$L(\beta, \lambda, \varsigma \text{ or } \delta) = \Pr[h_{11} < V_{11}, h_{21} < V_{21}, ..., h_{Q_1,1} < V_{Q_1,1}; \; h_{Q_1+1,1} > V_{Q_1+1,1}, h_{Q_1+2,1} > V_{Q_1+2,1}, ..., h_{Q,1} > V_{Q,1}], \qquad (12)$$

where $h_{q1} = u_{q2} - \varepsilon_{q1} = V_{q2} + (\varepsilon_{q2} - \varepsilon_{q1})$ for all $q = 1, 2, …, Q$. Even in this simple binary case, the likelihood function is very difficult to estimate, though Bhat and Sener (2009) and Bhat *et al.* (2010) have recently proposed computationally feasible and practical approaches to do so.[7] In the more general unordered case of Equation (11), the likelihood function is all but impossible to evaluate using simulation methods, because the h_{qi} terms are correlated across individuals for each *i*. In the current paper, we use a composite marginal likelihood (CML) approach that is gaining attention in the statistics field, though there has been little coverage of the method in econometrics and related fields (see Varin, 2008 and Bhat *et al.*, 2010 for recent reviews of this method).

2.2.1. The CML Approach

The CML approach is a useful inference approach when it is difficult or infeasible to evaluate the full information likelihood function, but when it is possible to compute the marginal likelihoods of subsets of the data. In this case, the analyst can form a pseudo-likelihood function by compounding the marginal likelihoods for the subsets. The resulting composite score function is a linear combination of legitimate likelihood score functions, and it is unbiased. This translates to the consistency and asymptotic normality of the CML estimator under usual regularity assumptions (Cox and Reid, 2004, Molenberghs and Verbeke, 2005, page 191). While there is a theoretical efficiency loss associated with the CML estimator relative to the full maximum likelihood estimator, this efficiency loss has been shown to be small in practice (see Henderson and Shimakura, 2003, and Lele, 2006).

In the current paper, we adopt a two-step CML procedure to develop a simple pseudo-likelihood function. First, we construct a first-level composite likelihood by taking the product of marginal likelihoods corresponding to the subset of observations that choose each of the alternatives

---

[7] As Bhat *et al.* (2010) indicate, extant "brute force" simulation methods within a classical or Bayesian framework such as those adopted by Bhat (2003), Beron and Vijverberg (2004), and LeSage (2000) are impractical and/or infeasible in binary choice situations with moderate sample sizes.

$i = 1, 2, \ldots, I-1$, and the marginal likelihood of the observations that choose the final alternative. Thus, the composite likelihood is written as:

$$L_{CML}^{*}(\beta, \lambda, \varsigma \text{ or } \delta) = \begin{bmatrix} \Pr(h_{11} < V_{11}, h_{21} < V_{21}, \ldots, h_{Q_1,1} < V_{Q_1,1}) \times \\ \Pr(h_{Q_1+1,2} < V_{Q_1+1,2}, h_{Q_1+2,2} < V_{Q_1+2,2}, \ldots, h_{Q_2,2} < V_{Q_2,2}) \times \\ . \\ . \\ . \\ \Pr(h_{Q_{(I-2)}+1,I-1} < V_{Q_{(I-2)}+1,I-1}, h_{Q_{(I-2)}+2,I-1} < V_{Q_{(I-2)}+2,I-1}, \ldots, h_{Q_{(I-1)},I-1} < V_{Q_{(I-1)},I-1}) \times \\ \Pr(D_{Q_{(I-1)}+1,I} = 1) \times \Pr(D_{Q_{(I-1)}+2,I} = 1) \times \ldots \Pr(D_{Q,I} = 1) \end{bmatrix} \tag{13}$$

The composite likelihood function above enables the explicit consideration of dependence across the h_{qi} terms for each alternative $i = 1, 2, \ldots, I-1$, which originates from the use of an appropriate copula. At the same time, it is very difficult to evaluate the joint probability of the second parenthesis term in Equation (11) for the final alternative, and thus we construct the composite marginal likelihood function using the marginal probability for each observation choosing the final alternative (note the independence in choice probabilities across observations in the final row of Equation 13).

2.2.2. The Pairwise Marginal Likelihood Inference Approach

The composite likelihood function of Equation (13) is still difficult to evaluate, especially for the Gaussian copula, which will entail the evaluation of $(I-1)$ multidimensional integrals (the first being of Q_1 dimensions, the second being of $Q_2 - Q_1$ dimensions, and so on). So, we further simplify the function in a second step to use a pairwise marginal likelihood estimation approach, which corresponds to a CML approach based on bivariate margins. In the process of doing so, we also introduce two weight terms ω_{qk} ($q, k \in \{1, 2, \ldots, Q\}; q \neq k$ and $\omega_{qk} = \omega_{kq}$) and μ_{qi} ($i = 1, 2, \ldots, I-1$), both of which we will define shortly. The CML function takes the following form:

$$L_{CML}(\beta, \lambda, \varsigma \text{ or } \delta) = \begin{bmatrix} \left(\prod_{q=1}^{Q_1-1} \left\{ \prod_{k=q+1}^{Q_1} \left[C_{\theta_{qk}}(u_{q1}, u_{k1}) \right]^{\omega_{qk}} \right\}^{\mu_{q1}} \right) \times \\ \left(\prod_{q=Q_1+1}^{Q_2-1} \left\{ \prod_{k=q+1}^{Q_2} \left[C_{\theta_{qk}}(u_{q2}, u_{k2}) \right]^{\omega_{qk}} \right\}^{\mu_{q2}} \right) \times \\ \vdots \\ \left(\prod_{q=Q_{(I-2)}+1}^{Q_{(I-1)}-1} \left\{ \prod_{k=q+1}^{Q_{(I-1)}} \left[C_{\theta_{qk}}(u_{q,I-1}, u_{k,I-1}) \right]^{\omega_{qk}} \right\}^{\mu_{q,I-1}} \right) \times \\ \left\{ \prod_{q=Q_{(I-1)}+1}^{Q} \frac{e^{V_{qI}/\sigma_q}}{\sum_{j=1}^{I} e^{V_{qj}/\sigma_q}} \right\} \end{bmatrix}$$

$$= \left[\prod_{i=1}^{I-1} \left(\prod_{q=Q_{i-1}+1}^{Q_i-1} \left\{ \prod_{k=q+1}^{Q_i} \left[C_{\theta_{qk}}(u_{qi}, u_{ki}) \right]^{\omega_{qk}} \right\}^{\mu_{qi}} \right) \right] \times \prod_{q=Q_{(I-1)}+1}^{Q} u_{qI}, \qquad (14)$$

where $u_{qi} = G_{qi}(V_{qi}) = \dfrac{e^{V_{qi}/\sigma_q}}{\sum_{j=1}^{I} e^{V_{qi}/\sigma_q}}$, $\sigma_q = \exp(\lambda' w_q)$, and the θ_{qk} terms are related to ς or δ as in

Equation (9) and Equation (10), respectively, for the FGM/Gaussian copula and the Generalized Gumbel copula ($Q_0 = 0$ and $Q_I = Q$ by notation in Equation 14). The *non-negative weight* terms, ω_{qk}, take the value of 1 if $k \in S_q$ and 0 otherwise, where S_q is the set of observational units k within a certain optimal threshold distance τ_{opt} of unit q.[8] The weights ω_{qk} are introduced because, in a spatial case where dependency drops quickly with inter-observation distance, the pairs formed from the closest observations provide much more information than pairs that are very far away. In fact, as demonstrated by Varin and Vidoni (2009) and Apanasovich *et al.* (2008) in different empirical contexts, retaining all observational pairs not only increases computational costs, but may also reduce estimator efficiency. The optimal distance, τ_{opt}, for inclusion of observation pairs may be

---

[8] In the empirical context of the current study. the distance between teenagers is computed as the Euclidean distance between the residence TAZ activity centroids of the teenagers.

set based on knowledge about the spatial process or based on minimizing the estimated asymptotic variance of estimators with varying values of the distance threshold (we will get back to this point at the end of the section). The *normalizing weight* terms, $\mu_{qi}(q \in \{Q_{i-1}+1, Q_i-1\}, i \in \{1, I\}$ in Equation (14), are inversely proportional to the number of pairings involving observation q (that chooses alternative i) with other observations choosing alternative i, and takes the specific form below:

$$\mu_{qi} = \frac{1}{\sum\limits_{\substack{k \neq q \\ k \in \{Q_{i-1}+1, Q_i\}}} \omega_{qk}}, q \in \{Q_{i-1}+1, Q_i\}, i = 1, 2, ..., I-1. \tag{15}$$

Defining M_i as the total number of observations choosing alternative i $(M_i = Q_i - Q_{i-1})$, it is easy to see that $\mu_{qi} = 1/(M_i - 1)$ when all pairings of individuals are considered within the group choosing each alternative i ($i = 1, 2, 3, ..., I - 1$). In this particular instance, and if there is no spatial correlation across individuals, it is straightforward to see that the composite likelihood function in Equation (14) collapses to the maximum likelihood function for an independent heteroscedastic multinomial logit (IHMNL) model.

The log composite likelihood function corresponding to Equation (14) is:

$$L_{CML}(\beta, \lambda, \varsigma \text{ or } \delta) = \log L_{CML}(\beta, \lambda, \varsigma \text{ or } \delta) = \left[\sum_{i=1}^{I-1} \sum_{q=Q_{i-1}+1}^{Q_i-1} \left[\mu_{qi} \sum_{k=q+1}^{Q_i} \omega_{qk} \ln C_{\theta_{qk}}(u_{qi}, u_{ki}) \right] \right] + \sum_{q=Q_{(I-1)}+1}^{Q} \ln u_{qI} \tag{16}$$

The above function can be maximized to obtain estimates of the relevant parameters. The CML estimator of $\gamma = (\beta', \lambda', \varsigma \text{ or } \delta)'$, obtained by maximizing the function in Equation (16) with respect to the vector γ, is asymptotically normal distributed with asymptotic mean γ and variance matrix given by the inverse of Godambe's (1960) sandwich information matrix (see Zhao and Joe, 2005, and Bhat *et al.*, 2010). To conserve on space, we do not provide details in the current paper for the estimator of the variance matrix, but the procedure is essentially the same as that in Bhat (2011).

The optimal threshold distance τ_{opt} that provides the most efficient parameter estimates in any copula model can be set by estimating the asymptotic variance matrix $Var(\gamma)$ of the estimator for different distance values and selecting the distance value that minimizes the total estimated variance across all parameters as given by $tr[Var(\gamma)]$, where $tr[A]$ denotes the trace of the matrix A.

3. DATA SOURCES AND SAMPLE FORMATION

In the current study, we examine the out-of-home weekday activity participation of teenagers in social, physically inactive recreational, and physically active recreational activities. The analysis is undertaken at an episode level, with the dependent variable being the type of activity (from among social, physically inactive recreation, and physically active recreation) participated in during each out-of-home episode.

3.1. The Primary Data Source

The primary source of data is the 2000 San Francisco Bay Area Travel Survey (BATS), which was designed and administered by MORPACE International, Inc. for the Bay Area Metropolitan Transportation Commission (see MORPACE International Inc., 2002). The survey collected detailed information on individual and household socio-demographic and employment-related characteristics from about 15,000 households in the Bay Area. The survey also collected information on all activity and travel episodes undertaken by individuals of the sampled households over a two-day period (the two day period comprised either two consecutive weekdays, or a Friday and a Saturday, or a Sunday and a Monday, but not a Saturday and a Sunday). The information collected on activity episodes included the type of activity (based on a 17-category classification system), the name of the activity participation location (for example, Jewish community center, Riverpark plaza, *etc.*), the type of participation location (such as religious place, or shopping mall), start and end times of activity participation, and the geographic location of activity participation.

3.2. The Secondary Data Source

In addition to the BATS survey, several other secondary Geographic Information System (GIS) data layers of highways, local roadways, bicycle facilities, businesses, and land-use/demographics were used to obtain spatial variables and neighborhood physical environment variables characterizing the residential traffic analysis zone (TAZ) of each teenager. The physical environment variables related to the residential neighborhood of teenagers include:

1) *Zonal land-use structure variables*, including housing type measures (fractions of single family, multiple family, duplex and other dwelling units), land-use composition measures (fractions of zonal area in residential, commercial, and other land-uses), and a land-use mix diversity index,

2) *Zonal size and density measures*, including total population, number of housing units, population density, household density, and employment density by several employment categories, as well as dummy variables indicating whether the area corresponds to a central business district (CBD), urban area, suburban area, or rural area.

3) *Regional accessibility measures*, which include Hansen-type (Fotheringham, 1983) employment, school, shopping, and recreational accessibility indices that are computed separately for the drive and transit modes.

4) *Zonal ethnic composition measures*, constructed as fractions of Caucasian, African-American, Hispanic, Asian and other ethnic populations for each zone.

5) *Zonal demographics and housing cost variables*, including average household size, median household income, and median housing cost in each zone.

6) *Zonal activity opportunity variables*, characterizing the composition of zones in terms of the intensity, the density or the presence of various types of activity centers. The typology used for activity centers includes five categories: (a) maintenance centers, such as grocery stores, gas stations, food stores, car wash, automotive businesses, banks, medical facilities, (b) physically active recreation centers, such as fitness centers, sports centers, dance and yoga studios, (c) physically passive recreational centers, such as theatres, amusement centers, and arcades, (d) natural recreational centers such as parks and gardens, and (e) restaurants and eat-out places. Note that the 'presence of an activity center' was defined by a dummy variable, which takes the value of one if there exist at least one (relevant) activity center, and zero otherwise.

7) *Zonal transportation network measures*, including highway and local roadway density (miles of highway facilities and local roadway facilities, respectively, per square mile), bikeway density (miles of bikeway facilities per square mile), street block density (number of blocks per square mile), and transit accessibility (number of zones connected by transit within 30 minutes).

8) *Spatial dependence variables*, characterizing the spatial dependence based on the residences of each pair of teenagers (these are the elements of the s_{qk} vector in Equation (9) and Equation (10) of Section 2.1). These include (1) whether or not two teenagers reside in the same TAZ, (2) whether or not two teenagers reside in contiguous TAZs, (3) the boundary length of the shared border between the residence zones of two teenagers, and (4) several functional forms of the

Euclidean distance ("crowfly" distance) between the residence TAZ activity centroids of the two teenagers, such as inverse of distance and square of inverse of distance.[9]

3.3. Sample Formation

The sample used for the current analysis is confined to a single weekday of 897 teenagers from 897 different households residing in nine Counties (Alameda, Contra Costa, San Francisco, San Mateo, Santa Clara, Solano, Napa, Sonoma, and Marin) of the San Francisco Bay Area. Since the empirical context of the paper is the social-recreational activity participation of teenagers, only individuals aged 13 to 19 years were considered in the analysis. Further, all activity episodes in which teenagers participated were classified by purpose, location (as out-of-home and in-home), and day of week, and only the weekday out-of-home social-recreational activity episodes were chosen for this study. The recreational activity episodes were further classified into physically active or physically inactive episodes, based on the activity type classification and the type of participation location.[10] That is, an episode designated as "recreation" activity by a respondent and pursued at a fitness center (such as working out at the gym) or pursued outdoors (such as walking/running/bicycling around the neighborhood "without any specific destination") is labeled as physically active. On the other hand, an episode designated as "recreation" activity by a respondent and pursued at a movie theater is labeled as physically inactive. The distribution of the number of social/recreational activity episodes in the sample is as follows: (1) social (30%), (2) physically inactive recreation (44.3%), and (3) physically active recreation (25.7%).

4. EMPIRICAL ANALYSIS

4.1. Model Specification

The model specification included variables falling into one of three broad categories:

(1) Individual characteristics, including age, sex, race, driver's license holding, and physical disability status.

---

[9] For two teenagers in the same zone, we assigned a distance that was one-half of the distance between that zone and its closest neighboring zone.

[10] A physically active episode requires regular bodily movement during the episode, while a physically passive episode involves maintaining a sedentary and stable position for the duration of the episode. For example, swimming or walking around the neighborhoods would be a physically active episode, while going to a movie is a physically inactive episode.

(2) Household characteristics, including number of adults, number of children, household composition and family structure, household income, dwelling type, whether the house is owned or rented, and parents' activity participation characteristics.

(3) Physical environment variables, including seasonal variables as well as the neighborhood physical environment variables related to the residential neighborhood of teenagers (as discussed in Section 3.2).

Table 1 presents the univariate descriptive statistics of each variable in the final model specification. The top row provides the total number of teenagers in the full sample (897 teenagers) as well as the number of teenagers participating in episodes of each of the three activity purpose categories. The remaining entries provide either the frequency (for categorical exogenous variables) or the mean value (for ordinal and continuous exogenous variables). Thus, the entry "454 (50.6)" for the "Female" variable in the "Full Sample" column indicates that 454 of the 897 teenagers are female, which corresponds to 50.6% of the teenagers. Similarly, the entry "155 (57.6)" for the "Female" variable in the "Social Activity" column indicates that 155 of the 269 teenagers who participate in social activities are female, which corresponds to 57.6% of the teenagers who participate in social activities. A comparison of the percentages across columns for categorical variables, and of the mean values across columns for ordinal/continuous variables, provides a preliminary sense of the directionality of the effects of variables. However, it should be kept in mind that these are but the univariate effect of each variable without controlling for the effects of other variables. The multivariate model results presented in Section 4.3 provides the more comprehensive picture. With that caveat, the statistics in Table 1 for the "female" variable indicate that female teenagers are more likely (relative to male teenagers) to participate in social activity and less likely to participate in active recreation, while teenagers of Hispanic origin are more likely (relative to teenagers of other races) to participate in social activity and less likely to participate in inactive recreation. Other observations may be similarly drawn.

A final summary statistic that may be of interest. The Euclidean distance (in miles) between the residence locations of teenagers varies from a minimum of 0.120 miles to a maximum of 151.46 miles with a mean value of 37.34 miles.

4.2. Model Selection

Three different nesting structure specifications of the three alternatives were considered to examine the possible presence of common unobserved effects in the utilities across alternatives for each teenager (for example, to test if a teenager who is more likely to participate in social activity episodes is also more likely to participate in physically inactive recreation episodes). These nesting structures considered in our specifications were either inconsistent with utility maximization or were not statistically superior to the simple multinomial logit model. Hence, we used the simple MNL model in the analysis. The optimal distance for selecting pairwise terms for inclusion in the composite likelihood was set based on minimizing the trace of the variance-covariance matrix of parameters. Specifically, we computed the trace of the variance matrix of parameters for various distance thresholds (that is, the threshold distance used to compute the ω_{qk} terms in Equation (15)), including 5 miles, 10 miles, 15 miles, 20 miles, 25 miles, 30 miles, 35 miles, 40 miles, 45 miles, 50 miles, 100 miles, and 151.46 miles, the last one representing the case of including all the $M_i(M_i-1)/2$ possible likelihood pairings of individuals choosing each alternative i in the CML function. Although the trace values did not change substantially based on the distance threshold used, the results showed that the best estimator efficiency, across all copula models, was obtained at about $\tau_{opt} = 45$ miles.

Among the three copula models (the FGM, the Gaussian, and the Generalized Gumbel), the Generalized Gumbel (GG) copula model (which we will refer to as the GGMNL model) turned out to be the preferred one based on Varin and Vidoni's (2005) composite likelihood information criterion. The implication in the current empirical context is that radially symmetric dependence patterns (such as those implied by the Gaussian and FGM copulas) may not be appropriate to capture spatial dependence in the types of activity episodes that teenagers participate in. Specifically, the dependence in unobserved components across teenagers in the propensity to participate in social-recreational episodes is strong at the right tail, but not at the left tail. That is, teenagers in close proximity (in terms of residence) tend to have uniformly high activity levels (tighter clustering of data points at the high end of the social-recreational utility spectrum), but there is lesser clustering of teenagers in close residential proximity toward the low activity levels (the low end of the social-recreational utility spectrum).

As we will show later in Section 4.4, the t-statistics on the heteroscedasticity and spatial dependence parameter estimates in the GGMNL model are highly statistically significant, indicating the statistically significant presence of heteroscedasticity and spatial effects. Furthermore, the results also show that the IHMNL and IMNL models provide less efficient estimates. In particular, the average of the trace of the covariance matrix of parameter estimates is 0.00136 for the GGMNL model, 0.00321 for the IHMNL model, and 0.00377 for the IMNL model, indicating the higher standard errors (by about 157%) from the IHMNL and IMNL models relative to the preferred GGMNL model. Overall, the recognition of spatial dependence (and heteroscedasticity) leads to substantially improved estimates of estimator efficiency. In fact, the results also indicate that the use of the IMNL model can potentially lead to inaccurate estimates regarding the effects of variables, as we discuss further in Section 4.5.

4.3. Estimation Results

To conserve on space, we only present the results for the best unordered response model (that is, GGMNL model). Table 2 presents the estimation results for this GGMNL model. The coefficients in the table provide the effects of variables on the latent propensity of teenagers to participate in social activities (first main column), physically inactive recreation activities (second main column), and physically active recreation activities (third main column). In instances where the coefficients on a variable for one or more alternatives are excluded, the omitted alternative category or categories form the base.

The coefficients on the alternative specific constants in Table 2 do not have substantive interpretations. They capture generic tendencies to participate in different activity alternatives as well as accommodate the range of continuous independent variables in the model.

4.3.1. Individual Characteristics

The effects of individual characteristics indicate that, among teenagers, females are less likely than males to participate in inactive and active recreation activities, and more likely than males to participate in social activities. Further, among recreational activities, the results show that females are particularly less likely to participate in physically active recreation compared to their male peers. These results are consistent with several previous studies, including King *et al.*, 2007 and Trolano *et al.*, 2008. For instance, King *et al.* (2007) found that girls are likely to participate more intensively,

and have a higher enjoyment, in social and self-improvement activities, while boys are more likely to participate, and have a higher enjoyment, in physically active recreation activities. Of course, the reasons for such sex-related differences need to be studied in much more depth to understand the influence of environmental and societal expectations/norms on such stereotypical inclinations.

The race-related effects suggest that Hispanic teenagers are less predisposed (relative to their peers of Caucasian, African, Asian, and other racial descents) to engage in inactive recreation, while Asian teenagers are more predisposed toward social activities. Part-time students have a higher propensity (relative to full-time students and non-students) to participate in recreational activities rather than social activities. However, the above race and part-time student effects should be viewed with some caution because of the relatively small sizes of these groups in the sample. Finally, the results indicate that teenagers with a driver's license are more likely to participate in social activities than in recreational activities. This is perhaps a reflection of the freedom to drive to social activity opportunities that may be far away from one's own residential neighborhood (such as a party at a friend's place who lives far away). However, we also noticed that this variable serves as a proxy for age-related effects (in fact, when we introduced a dummy variable for age greater than 16 years, the statistical significance of the driver's license coefficient dropped; however, because of multicollinearity problems, the age-related effect was not statistically significant). Earlier studies have also suggested that, as children get older, they gravitate more toward unstructured social activities rather than structured sports activities and unstructured free play (see Sallis *et al*., 2000, and Copperman and Bhat, 2007).

4.3.2. Household Characteristics

The household-related variable effects show the higher propensity among teenagers living in nuclear family households (*i.e*., households with both parents living with the teenager) to participate in out-of-home active and inactive recreational activities rather than social activities. This may be a reflection of the increased time availability of adults in nuclear family households to engage in joint recreational activities with children, though the result needs further exploration (for instance, the result may also be a reflection of teenagers with divorced parents spending time visiting the parent with whom they do not live, leading to the higher social activity participation and reduced recreation participation of such teenagers compared to teenagers in nuclear families).

The results corresponding to the household income variable point to the positive effect of higher socio-economic status on social and inactive recreational activity participation. This may be due to financial constraints in low income households, which hamper the ability to partake in social/inactive recreation (for example, going to the movies entails an admission fee). However, the results do suggest that income does not appear to be an economic factor in participation in active recreation. This is somewhat surprising, given that earlier transportation and public health studies (see Loukaitou-Sideris, 2004 and Day, 2006) have indicated that adults from low income households partake less in active recreation, both in terms of walking/bicycling around their neighborhoods (potentially because of poor non-motorized mode infrastructure and safety/security concerns) as well as in terms of physical activity at gyms and health clubs (potentially because of the financial cost). Whether the difference in the finding in this paper and those of earlier papers is because of the segment of the population studied (teenagers as opposed to adults), or due to our focus on the San Francisco Bay area with its relatively rich mixed land-use and non-motorized mode infrastructure (that may provide better opportunities for inexpensive active recreation opportunities such as walking and bicycling around the neighborhood), or some other factors is an issue that needs further exploration.

The effect of the number of motorized vehicles in Table 2 indicates a tendency of teenagers from households with several motorized vehicles to participate more in inactive recreation than in social or active recreation episodes. Finally, in the category of household variables, the physical activity levels of parents positively influences the physical activity levels of teenagers (for the purpose of this research, we designate a parent as participating in physically active recreation if the parent pursues one or more active recreation episodes on the survey day). As emphasized in the literature (see, for instance, see Davis *et al.*, 2007 and Sener *et al.*, 2010), this is because teenagers (and children) explicitly model their parent's physical activity participation (or physically active lifestyle). Further, the joint recreational activity participation of parents and teenagers can significantly lead to an increased level of physical activity for both teenagers and their parents. These results suggest the importance of family-oriented educational campaigns to increase awareness regarding the benefits of a physically active lifestyle. For instance, middle school and high schools may want to consider organizing information sessions on health and physical activity for parents of students in their schools, rather than confining health-related instruction to students.

4.3.3. Physical Environment Variables

The first variable among the physical environment variables corresponds to seasonal effects. As one would expect, teenagers are less likely to participate in physically active recreational pursuits during the winter months, presumably because the weather conditions may discourage outdoor physical activity during the cold winter months relative to the other times of the year (see also Tucker and Gilliland, 2007 and Sener *et al.*, 2010 for a similar finding).[11] The lower physical activity participation of teenagers in active out-of-home activities during winter should be carefully examined by policy makers to increase (and balance) physical activity participation throughout the year. One possible consideration is to provide more indoor activity opportunity centers at affordable cost and close to residential neighborhoods.

The next set of variables highlights the importance of the residential location and built environment variables. Of course, the effects of this set of variables should be viewed with some caution since we have not considered potential residential self-selection effects. That is, it is possible that highly physically active families self-select themselves into zones with built environment measures that support their active lifestyles (see Bhat and Guo, 2007, and Bhat and Eluru, 2009 for methodologies to accommodate such self-selection effects). With that caveat, the results show that accessibility to schools has a significant and positive effect on recreational activity participation (both physically inactive and active). This is probably a reflection of the location flexibility motivating and increasing teenager's recreational activity participation. That is, teenagers going to schools within close proximity of their residences are less dependent on their parents (or on other adult household members) for transportation to/from schools, and can walk/bicycle to school for recreational activity participation. Also, as indicated recently by Paleti *et al.* (2011) in their extensive examination of children's non-school activity participation based on the 2002 Child Development Supplement (CDS) of the Panel Study of Income Dynamics (PSID), a unique aspect of children's activity-travel patterns is the role school plays as a significant location for out-of-home organized and recreation activity participation. Thus, it is not at all surprising that school accessibility has a strong positive influence on recreation activity participation. One should also note

---

[11] Admittedly, the winter weather conditions in San Francisco are not that harsh from an absolute temperature standpoint as in other northern parts of the country such as Wisconsin or North Dakota. However, winter months are still colder in San Francisco relative to the other times of the year. Given that human beings tend to adapt themselves to the conditions they live in, an individual residing in San Francisco will therefore perceive the winter months as being cold compared to the other parts of the year.

that school accessibility, while promoting the physically active/inactive recreational activity participation at school, can also lead to an increase in the physical activity levels of children due to active transportation (though this is not a subject of our study). Next, the presence of physical activity opportunity centers in a zone increases the active recreation among teenagers residing in the zone, suggesting the importance of providing more such centers in close proximity of residences and/or developing attractive, accessible and affordable physically active recreation programs at neighborhood youth community centers. Teenagers living in residential areas with a high bicycle facility density (as measured by miles of bicycle lanes per square mile in the residential TAZ) have a higher likelihood to partake in physically active recreation and social activities rather than physically inactive recreational activities, suggesting the potential benefits of dense, mixed land-use, walkable and bikeable neighborhoods for the promotion of socially vibrant and physically active life styles (see also Krizek *et al.*, 2004, and Bhat and Sener, 2009). The results also indicate the increased likelihood to participate in physically inactive recreation with good transit accessibility.

Finally, teenagers residing in dense and urban environments are more likely to engage in social activities, and less likely to pursue recreational activities, relative to teenagers in less dense and non-urban environments. This result needs to be explored and acted upon further, since a healthy balance in both social activity participation (to build identity and healthy relationships with peers) and physically active recreation (to enhance mental and physical state of the mind and body) is important for the development of teenagers.

4.4. Heteroscedasticity and Spatial Dependency

This section presents the parameter estimates characterizing heteroscedasticity and spatial dependence in the teenager's social-recreational activity participation model.

4.4.1. Heteroscedasticity

As discussed in Section 2.1, the model formulated in the current paper allows the incorporation of heteroscedasticity among individuals through the λ vector embedded in $\sigma_q = \exp(\lambda' w_q)$. We examined the effect of several variables in the w_q vector, and those that turned out to be statistically significant are presented in Table 2 under the label of "(Spatial) heteroscedasticity variables" corresponding to the λ vector. The results indicate a tighter variation (*i.e.*, less spread) in the social-

recreational activity propensity of teenagers in nuclear families relative to teenagers in other family types. This effect, in conjunction with the direct positive effect of nuclear families on recreational activity participation, suggests a uniformly higher propensity of teenagers in nuclear family households to participate in recreational activities. Further, the results also indicate a much tighter variation in the propensity of social-recreational activity participation on Fridays compared to other days of the week, and among teenagers residing in zones with out-of-home recreational activity centers relative to teenagers in zones with no out-of-home recreational activity centers.

4.4.2. Spatial Dependence Effects

As indicated in Section 4.2, the GGMNL model provided the best fit to capture spatial dependence effects. The dependency among observational units in this model is captured through the dependence vector θ, including the ρ term as well as the $\alpha_{q,qk}$ terms (see Section 2.1). The estimated value of the ρ parameter is 0.570, with a standard error of 0.0251. A t-statistic test with respect to 1 (which represents the case of no spatial dependence) returns a value of 17.13, which is higher than the t-statistic table value at any reasonable level of significance. The second parameter under the "Spatial dependence variables" corresponds to the $\widetilde{s}_{qk}$ vector (and the corresponding δ coefficient vector) of $\alpha_{q,qk}$ in Equation (10). The best specification for the unordered model included a single "inverse of distance variable" (distance being measured as the spatial separation in miles between the residence TAZ centroids of teenagers) in the s_{qk} vector. In other words, as the distance between a teenager's residence zone and another teenager's residence zone increases, the degree of dependency in the propensities to pursue social activity episodes decreases. The same holds true for the physically inactive and physically active recreation purposes.

The above discussion highlights that the spatial dependence effect is very highly statistically significant, and needs to be accommodated. The IMNL model completely ignores these spatial dependencies, while the proposed copula model (GGMNL) explicitly considers both spatial dependency and (spatial) heteroscedasticity. The result, as indicated earlier, is that the IMNL model provides less efficient estimates, which can have implications regarding inferences associated with the effects of variables. Further, as we discuss in the next section, the IMNL model also provides inconsistent elasticity effects.

4.5 Aggregate-Level Elasticity Effects

The parameters on the exogenous variables in Table 2 do not directly provide the magnitude of the effects of variables in the choice probabilities of each episode type. To address this issue, we compute the aggregate-level "elasticity effects" of variables.

To compute an aggregate-level "elasticity" of an ordinal exogenous variable (such as the number of household vehicles), we increase the value of the ordinal variable by 1 unit for each individual and obtain the relative change in expected aggregate shares. Thus, the "elasticities" for the ordinal exogenous variables can be viewed as the relative change in expected aggregate shares due to an increase of 1 unit in the ordinal variable across all individuals. To compute an aggregate-level "elasticity" of a dummy exogenous variable (such as whether a teenager is female, is in a nuclear family, or is Asian), we change the value of the variable to one for the subsample of observations for which the variable takes a value of zero and to zero for the subsample of observations for which the variable takes a value of one. We then sum the shifts in expected aggregate shares in the two subsamples after reversing the sign of the shifts in the second subsample and compute an effective proportional change in expected aggregate shares in the entire sample due to a change in the dummy variable from 0 to 1. Finally, the aggregate-level "arc" elasticity effect of a continuous exogenous variable (such as bicycle facility density) is obtained by increasing the value of the corresponding variable by 10% for each individual in the sample, and computing a percentage change in the expected aggregate shares of each activity type. While the aggregate level elasticity effects are not strictly comparable across the three different types of independent variables (dummy, ordinal, and continuous), they do provide order of magnitude effects.

The elasticity effects by variable category for both the (aspatial) IMNL model and the proposed GGMNL model are presented in Table 3. The results reveal that parent's physical activity participation constitutes the most important factor positively influencing teenagers' physical activity participation level. This result highlights the importance of increasing the awareness of the health benefits of active recreation among parents, which would have a direct influence on teenager's physical activity participation. In addition, being a part-time student is also found to have a substantial positive impact on teenagers' recreational activity participation levels (both physically active and physically inactive). This may be because part-time students, relative to their full-time peers, may have more time to spend in after-school activities (such as music courses, arts-crafts, or coached sports) pursued at (school-) clubs on weekdays. If so, the results suggest that an effective

component of a successful PYD program, as discussed in Section 3.1., would be to provide more opportunities for teenagers to participate in flexible, yet organized, recreational activities to help them discover their talents, strengths and full potential. Of course, the findings and recommendations with regard to part-time students should be viewed with caution, given the small fraction of teenagers in the sample who are part-time students. Another important result from the table is the strong gender differences in different types of activity participation. Specifically, the results show that females are much more likely than males to participate in social activities, and much less likely to partake in active recreation activities. These gender stereotype differences require careful further examination. Furthermore, the results also indicate the lower likelihood of active recreation during the winter season compared to other seasons, perhaps emphasizing the importance of public health policies aimed at providing (more) indoor active recreation opportunities (although this lower winter season physical activity pattern may also simply be a reflection of the sluggish winter "mood" of individuals and/or related to the holiday period associated with the winter season). Again, the finding and recommendation regarding seasons should be viewed with some caution given the small sample size of households in the sample interviewed in the winter season.

There are also differences in the elasticity effects between the aspatial IMNL model and the spatial GGMNL model. This, in conjunction with the better data fit of the spatial model, points to the inconsistent elasticity effects from the IMNL model. For instance, while the IMNL model indicates a positive effect of being in a nuclear family on teenagers' active recreation activity, the GGMNL model indicates a mild negative effect on active recreation activity. There are also quite substantial differences in the effects of the physical environment variables between the aspatial and spatial models (see, for example, the differential effects of the "winter" dummy variable, the "presence of physically active recreation centers", the "logarithm of household population density in zone", and the "urban" residence variables). It is also important to note that the IMNL model fails to capture the impact of two of the variables (the "presence of out-of-home recreational activity centers" variable and the "Friday" dummy variable) on social-recreational activity participation, while these variables have relatively high effects in the GGMNL model, especially on the active recreation category. Overall, ignoring spatial effects, when present, and using the IMNL model, can lead to inconsistent parameter estimates, which may result in unreliable policy evaluations.

5. SUMMARY AND CONCLUSIONS

The current paper proposes an approach to accommodate flexible spatial dependency structures in discrete choice models in general, and in unordered multinomial choice models in particular. Specifically, we combine a copula-based formulation for spatial dependence in an unordered multinomial response model with a pseudo-likelihood estimation technique based on a composite marginal likelihood (CML) inference approach. While the copula approach provides a flexible structure for incorporating spatial dependence (that do not impose any restrictive assumption on the dependency structure), the proposed CML estimation approach leads to a simple and practical approach, which is applicable to data sets of any size and does not require any simulation machinery.

The proposed copula-CML model is applied to examine teenagers' participation in social-recreational activity purposes, a subject of considerable interest in the adolescence development, public health, and transportation fields. The data for the analysis is drawn from the 2000 San Francisco Bay Area Survey. A flexible spatial error dependence in participation propensities in the activity purposes is generated across teenagers based on the proximity of their residences. Such dependencies may be the result of unobserved residential urban form effects and/or diffusion and social interaction effects between proximally located teenagers, so that unobserved lifestyle perspectives (such as physically active lifestyle attitudes) that affect activity participation decisions become correlated. Several copula model forms were tested to capture the spatial error dependencies across teenagers during the empirical specification, from which the Generalized Gumbel (GG) copula formulation emerged as the best specification (that is, provided the best data fit). This implies that teenagers in close proximity (in terms of residence) tend to have uniformly high activity levels (tighter clustering of data points at the high end of the social-recreational utility spectrum), but there is lesser clustering of teenagers in close residential proximity toward the low activity levels (low end of the social-recreational utility spectrum). The variable effects indicate that parents' physical activity participation constitutes the most important factor influencing teenagers' physical activity participation levels, suggesting that one of the most effective ways to increase active recreation among teenagers would be to direct physical activity benefit-related information and education campaigns toward parents, perhaps at special physical education sessions at schools for parents of teenagers studying there. In addition, part-time student status, gender, and seasonal effects are also important determinants of teenagers' social-recreational activity participation.

ACKNOWLEDGEMENTS

REFERENCES

Apanasovich T.V., Ruppert D., Lupton J.R., Popovic N., Turner N.D., Chapkin R.S., Carroll R.J. (2008) Aberrant crypt foci and semiparametric modeling of correlated binary data. *Biometrics* 64(2): 490-500.

Beron K.J., Vijverberg W.P.M. (2004) Probit in a spatial context: a Monte Carlo analysis. In: Anselin L., Florax R.J.G.M., Rey S.J. (eds), *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Springer-Verlag, Berlin. pgs. 169-196.

Bhat C.R. (2003) Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transportation Research Part B* 37(9): 837-855.

Bhat C.R. (2009) A new generalized Gumbel copula for multivariate distributions. Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, August 2009.

Bhat C.R. (2011) The maximum approximate composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. *Transportation Research Part B* 45(7): 923-939.

Bhat C.R., Eluru N. (2009) A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. *Transportation Research Part B* 43(7): 749-765.

Bhat C.R., Guo J.Y. (2007) A comprehensive analysis of built environment characteristics on household residential choice and auto ownership levels. *Transportation Research Part B* 41(5): 506-526.

Bhat C.R., Sener I.N. (2009) A copula-based closed-form binary logit choice model for accommodating spatial correlation across observational units. *Journal of Geographical Systems* 11(3): 243-272.

Bhat C.R., Sener I.N., Eluru N. (2010) A flexible spatially dependent discrete choice model: formulation and application to teenagers' weekday recreational activity participation. *Transportation Research Part B* 44(8-9): 903-921.

Boarnet M.G., Chalermpong S., Geho E. (2005) Specification issues in models of population and employment growth. *Papers in Regional Science* 84(1): 21-46.

Campbell J. (2007) Adolescent identity development: the relationship with leisure lifestyle and motivation. Master of Arts Thesis, Department of Recreation and Leisure Studies, University of Waterloo, Waterloo, Ontario, Canada.

Centers for Disease Control and Prevention (CDC) (2005) Positive parenting tips for healthy child development. Department of Health and Human Services, National Center on Birth Defects and Developmental Disabilities. Online at http://www.cdc.gov/ncbddd/child/documents/6-8YearOldsPositiveParenting.pdf

Centers for Disease Control and Prevention (CDC) (2006) Youth risk behavior surveillance – United States, 2005. Morbidity and Mortality Weekly Report Surveillance Summaries, 55(SS-5).

Centers for Disease Control and Prevention (CDC) (2010) Youth risk behavior surveillance – United States, 2009. Morbidity and Mortality Weekly Report Surveillance Summaries, 59(SS-5).

Cho W.T., Rudolph T. (2007) Emanating political participation: untangling the spatial structure behind participation. *British Journal of Political Science* 38(2): 273-289.

Copperman R.B., Bhat C.R. (2007) An exploratory analysis of children's daily time-use and activity patterns using the child development supplement (CDS) to the US panel study of income dynamics (PSID). *Transportation Research Record* 2021: 36-44.

Cox D., Reid N. (2004) A note on pseudolikelihood constructed from marginal densities. *Biometrika* 91(3): 729-737.

Darling N. (2005) Participation in extracurricular activities and adolescent adjustment: cross-sectional and longitudinal findings. *Journal of Youth and Adolescence* 34(5): 493-505.

Davis M.M., Gance-Cleveland B., Hassink S., Johnson R., Paradis G., Resnicow K. (2007) Recommendations for prevention of childhood obesity. *Pediatrics* 120(suppl4): S229-S253.

Day K. (2006) Active living and social justice: planning for physical activity in low-income, black, and latino communities. *Journal of American Planning Association* 72(1): 88-99.

Eccles J.S. (1999) The development of children ages 6 to 14. *The Future of Children* 9(2): 30-44.

Eccles J., Gootman J.A. (2002) *Community Programs to Promote Youth Development*. Committee on Community-Level Programs for Youth, Board on Children, Youth, and Families, Commission on Behavioral and Social Sciences Education, National Academy Press, Washington, DC.

Fotheringham A.S. (1983) Some theoretical aspects of destination choice and their relevance to production-constrained gravity models. *Environment and Planning* 15(8): 1121-1132.

Franzese R.J., Hays J.C. (2008) Empirical models of spatial interdependence. In: Box-Steffensmeier J., Brady H., Collier D. (eds), *Oxford Handbook of Political Methodology*, Oxford University Press, pp. 570-604.

Fredricks J.A., Eccles J.S. (2008) Participation in extracurricular activities in the middle school years: are there developmental benefits for African American and European American youth? *Journal of Youth and Adolescence* 37(9): 1029-1043.

Godambe V. (1960) An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics* 31(4): 1208-1211.

Henderson R., Shimakura S. (2003) A serially correlated gamma frailty model for longitudinal count data. *Biometrika* 90(2): 355-366.

Jones K., Bullen N. (1994) Contextual models of urban home prices: a comparison of fixed and random coefficient models developed by expansion. *Economic Geography* 70(3): 252-272.

King G., Law M., King S., Hurley P., Hanna S., Kertoy M., Rosenbaum P. (2007) Measuring children's participation in recreation and leisure activities: construct validation of the CAPE and PAC. *Child: Care, Health and Development* 33(1): 28-39.

Krizek K., Birnbaum A., Levinson D. (2004) A schematic for focusing on youth in investigation of community design and physical activity. *American Journal of Health Promotion* 19(1): 33-38.

Larson R.W. (2000) Toward a psychology of positive youth development. *American Psychologists* 55(1): 170-183.

Lele S.R. (2006) Sampling variability and estimates of density dependence: a composite-likelihood approach. *Ecology* 87(1): 189-202.

Lerner R., Steinberg L. (2004). *Handbook of Adolescent Psychology* (2nd ed.), Wiley, New York.

LeSage J.P. (2000) Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis* 32(1): 19-35.

Loukaitou-Sideris A. (2004) Transportation, land use, and physical activity - safety and security considerations. *TRB Special Report 282 – Does the Built Environment Influence Physical Activity? Examining the Evidence*. Paper prepared for the Transportation Research Board and the Institute of Medicine Committee on Physical Activity, Health, Transportation, and Land Use.

McDonald N. (2005) Children's travel: patterns and influences. Ph.D. Dissertation, University of California, Berkeley.

McGuckin N., Nakamoto Y. (2004) Differences in trip chaining by men and women. Research on Woman's Issues in Transportation: Report of a Conference, Vol. 2: Technical Papers, Transportation Research Board, Nov. 18-20, 2004. Chicago, IL.

Messner S.F., Anselin L. (2004) Spatial analyses of homicide with areal data. In: Goodchild M., Janelle D. (eds), *Spatially Integrated Social Science*, Oxford University Press, New York, pgs 127-144.

Miller H.J. (1999) Potential contributions of spatial analysis to geographic information systems for transportation (GIS-T). *Geographical Analysis* 31(4): 373-399.

Molenberghs G., Verbeke G. (2005) *Models for Discrete Longitudinal Data*. Springer Science + Business Media, Inc., New York.

MORPACE International, Inc. (2002) Bay Area Travel Survey Final Report. Metropolitan Transportation Commission, CA. Available at:
ftp://ftp.abag.ca.gov/pub/mtc/planning/BATS/BATS2000/

Nelson M.C., Gordon-Larsen P. (2006) Physical activity and sedentary behavior patterns are associated with selected adolescent health risk behaviors. *Pediatrics* 117(4): 1281-1290.

Ornelas I.J., Perreira K.M., Ayala G.X. (2007) Parental influences on adolescent activity: a longitudinal study. *The International Journal of Behavioral Nutrition and Physical Activity* 4: 3-10.

Páez A. (2007) Spatial perspectives on urban systems: developments and directions. *Journal of Geographic Systems* 9(1): 1-6.

Paleti R., Copperman R.B., Bhat C.R. (2011) An empirical analysis of children's after school out-of-home activity-location engagement patterns and time allocation. *Transportation* 38(2): 273-304.

Reisner E. (2003) Understanding family travel demands as a critical component in work-family research, transportation and land-use. Presented at *From 9 to 5 to 24/7: How Workplace Changes Impact Families, Work and Communities, Academic Work and Family Research Conference*, March.

Sallis J.F., Prochaska J.J., Taylor W.C. (2000) A review of correlates of physical activity of children and adolescents. *Medicine and Science in Sports and Exercise* 32(5): 963-975.

Sanchez-Samper X., Knight J.R. (2009) Drug abuse by adolescents. *Pediatrics in Review* 30(3): 83-93.

Sener I.N., Bhat C.R. (2007) An analysis of the social context of children's weekend discretionary activity participation. *Transportation* 34(6): 697-721.

Sener I.N., Copperman R.B., Pendyala R.M., Bhat C.R. (2008) An analysis of children's leisure activity engagement: examining the day of week, location, physical activity level, and fixity dimensions. *Transportation* 35(5): 673-696.

Sener I.N., Eluru N., Bhat C.R. (2010) On jointly analyzing the physical activity participation levels of individuals in a family unit using a multivariate copula framework. *Journal of Choice Modelling* 3(3): 1-38.

Sklar A. (1973) Random variables, joint distribution functions, and copulas. *Kybernetika* 9: 449-460.

Stefan K.J., Hunt J.D. (2006) Age-based analysis of children in Calgary, Canada. Presented at the 85th Annual Meeting of the Transportation Research Board, Washington D.C., January.

Tiggemann M. (2001) The impact of adolescent girls' life concerns and leisure activities on body dissatisfaction, disordered eating and self-esteem. *The Journal of Genetic Psychology* 162(2): 133-142.

Trolano R.P., Berrigan D., Dodd K., Masse L.C., Tilert T., McDowell M. (2008) Physical activity in the United States measures by accelerometer. *Medicine & Science in Sports & Exercise* 40(1): 181-188.

Tucker P., Gilliland J. (2007) The effect of season and weather on physical activity: a systematic review. *Public Health* 121: 909-922.

Varin C. (2008) On composite marginal likelihoods. *AStA Advances in Statistical Analysis* 92(1): 1-28.

Varin C., Vidoni P. (2005) A note on composite likelihood inference and model selection. *Biometrika* 92(3): 519-528.

Varin C., Vidoni P. (2009) Pairwise likelihood inference for general state space models. *Econometric Reviews* 28(1-3): 170-185.

Zhao Y., Joe H. (2005) Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics* 33(3): 335-356.

TABLE 1 Descriptive Statistics of the Sample Data

	Value "#(%) or Mean"			
		Teenager's participating in...*		
Variable	**Full Sample (n=897)**	**Social Activity (n=269)**	**Inactive Recreation Activity (n=397)**	**Active Recreation Activity (n=231)**
Individual characteristics				
Female	454 (50.6)	155 (57.6)	200 (50.4)	99 (42.9)
Hispanic	48 (5.4)	19 (7.1)	16 (4.0)	13 (5.6)
Asian	56 (6.2)	20 (7.4)	24 (6.0)	12 (5.2)
Part-time student	24 (2.7)	4 (1.5)	12 (3.0)	8 (3.5)
Licensed driver	365 (40.7)	139 (51.7)	142 (35.8)	84 (36.4)
Household characteristics				
Nuclear family	423 (47.2)	107 (39.8)	201 (50.6)	115 (49.8)
Household income greater than 90K	367 (40.9)	116 (43.1)	173 (43.6)	78 (33.8)
Number of household vehicles	2.67	2.69	2.98	2.64
Teenager's mother physically active	204 (22.7)	47 (17.5)	71 (17.9)	86 (37.2)
Teenager's father physically active	154 (17.2)	30 (11.2)	64 (16.1)	60 (26.0)
Activity-day variables				
Friday	169 (18.8)	56 (20.8)	77 (19.4)	36 (15.6)
Physical environment variables				
Winter	62 (6.9)	17 (6.3)	36 (9.1)	9 (3.9)
Residence location/neighborhood variables				
Accessibility to schools	0.08033	0.07997	0.08097	0.07964
Presence of physically active recreation centers	588 (65.6)	186 (69.1)	243 (61.2)	159 (68.8)
Presence of out-of-home recreational activity centers	771 (86.0)	241 (89.6)	337 (84.9)	193 (83.5)
Bicycle facility density (miles of bike lanes per square mile)	1.83	2.06	1.65	1.85
Number of zones connected by transit within 30 minutes	7.80	6.80	8.35	8.01
Logarithm of household population density	8.72	8.77	8.72	8.67
Urban residence	130 (14.5)	46 (17.1)	53 (13.4)	31 (13.4)

* Percentages are based on total number of teenagers participating in each activity type.

TABLE 2 GGMNL Model Estimation Results for the Teenager's Social-Recreational Activity Participation

Variable	Social		Inactive Recreation		Active Recreation	
	Parameter	t-stat	Parameter	t-stat	Parameter	t-stat
Alternative specific constants	-	-	0.362	0.86	0.207	0.55
Individual characteristics						
Female	-	-	-0.083	-2.32	-0.169	-2.83
Hispanic	-	-	-0.181	-2.98	-	-
Asian	0.140	3.24	-	-	-	-
Part-time student	-	-	0.363	3.08	0.363	3.08
Licensed driver	0.139	2.95	-	-	-	-
Household characteristics						
Nuclear family	-	-	0.085	1.13	0.085	1.13
Household income greater than 90K	0.093	2.24	0.093	2.24	-	-
Number of household vehicles	-	-	0.019	1.94	-	-
Teenager's mother physically active	-	-	-	-	0.210	2.79
Teenager's father physically active	-	-	-	-	0.133	2.54
Physical environment variables						
Seasonal variables	-	-	-	-	-	-
Winter	-	-	-	-	-0.238	-2.78
Residential location and built environment variables						
Accessibility to schools	-	-	1.993	2.66	1.993	2.66
Presence of physically active recreation centers	-	-	-	-	0.097	2.59
Bicycle facility density (miles of bike lanes per square mile)	0.022	2.69	-	-	0.013	1.80
Number of zones connected by transit within 30 minutes	-	-	0.005	2.79	-	-
Residential neighborhood variables						
Logarithm of household population density in zone	0.054	2.13	-	-	-	-
Urban	0.160	2.60	-	-	-	-
(Spatial) heteroscedasticity variables						
Nuclear family	-0.394	-3.17	-0.394	-3.17	-0.394	-3.17
Friday	-0.408	-2.36	-0.408	-2.36	-0.408	-2.36
Presence of out-of-home recreational activity centers	-1.206	-3.25	-1.206	-3.25	-1.206	-3.25
Spatial dependence variables						
ρ term *	-	-	0.570	17.13	0.570	17.13
δ in the α parameter "*Inverse of distance between zonal centroids*"	-	-	0.797	2.33	0.797	2.33
Number of Observations	897					
Trace of G	1.232					
Log-composite likelihood at convergence	-693.966					
Trace of the matrix in the CLIC statistic	142.352					
Penalized log-composite likelihood (PLCL)	-836.318					

* The t-statistic is computed for the null hypothesis that $\rho = 1$

TABLE 3 Aggregate-level Elasticity Effects of the Aspatial IMNL and Spatial GGMNL Model

Variable	Aspatial IMNL Model			Spatial GGMNL Model		
	Social	Inactive Recreation	Active Recreation	Social	Inactive Recreation	Active Recreation
Individual characteristics						
Female	25.88	-4.87	-35.29	25.08	-5.16	-34.72
Hispanic	19.01	-29.74	19.77	24.26	-38.22	23.61
Asian	27.85	-17.53	-16.52	38.18	-21.46	-19.00
Part-time student	-52.27	33.26	30.37	-61.89	41.62	35.95
Licensed driver	37.80	-23.82	-22.36	31.85	-21.33	-18.65
Household characteristics						
Nuclear family	-17.46	11.03	10.30	-14.55	15.84	-1.96
Household income greater than 90K	8.27	9.22	-30.10	6.83	7.70	-25.13
Number of household vehicles	-1.93	3.01	-1.99	-2.28	4.46	-2.76
Teenager's mother physically active	-18.56	-20.37	66.98	-17.58	-19.64	64.36
Teenager's father physically active	-12.07	-13.27	43.61	-10.94	-12.17	39.96
Activity-day variables						
Friday	-	-	-	4.82	3.32	-14.09
Physical environment variables						
Winter	11.73	13.21	-42.92	14.01	15.76	-51.48
Accessibility to schools	-3.37	2.39	2.14	-3.39	2.32	1.90
Presence of physically active recreation centers	-2.64	-2.93	9.58	-6.87	-7.74	25.28
Presence of out-of-home recreational activity centers	-	-	-	10.47	9.14	-33.93
Bicycle facility density (miles of bike lanes per square mile)	0.87	-1.02	0.32	0.89	-0.97	0.12
Number of zones connected by transit within 30 minutes	-0.50	0.78	-0.52	-0.74	1.17	-0.72
Logarithm of household population density in zone	17.72	-11.17	-10.47	12.19	-27.82	-26.70
Urban	35.89	-22.58	-35.89	36.37	-24.22	-22.55